Parallel-NetCDF: A High Performance API for NetCDF File Access

Overview

Parallel-NetCDF is a library providing high-performance I/O while still maintaining file-format compatibility with Unidata's NetCDF.

NetCDF gives scientific programmers a space-efficient and portable means for storing data. However, it does so in a serial manner, making it difficult to achieve high I/O performance. By making some small changes to the API specified by NetCDF, we can use MPI-IO and its collective operations.

- <u>Download</u> has the latest release and development links as well as information about svn access.
- <u>Documentation</u>: a <u>QuickTutorial</u>, plus papers, presentations, articles, and other resources
- <u>Benchmarking</u>: tools and suggestions for evaluating pnetcdf performance

News

- Nonblocking I/O is redesigned in the 1.2.0 release. It defers the I/O requests until "wait" call, so small requests can be aggregated into large ones for better performance.
- Two new hints, `nc_header_align_size` and `nc_var_align_size`, are added. The former allows pre-allocation of a larger header size to accommodate new header data in case new variables or attributed are added later. The latter aligns the starting file offsets of non-record variables. Refer to <u>VariableAlignment</u> for a more detailed description.
- Data consistency control has been revised. A more strict consistency can be enforced by using `NC_SHARE` mode at the file open/create time. In this mode, the file header is synchronized to the file if its contents have changed. Such file synchronization of calling `MPI_File_sync()` happens in many places, including `ncmpi_enddef()`, `ncmpi_redef()`, all APIs that change global or variable attributes, dimensions, and number of records.
- As calling `MPI_File_sync()` is very expensive on many file systems, users can choose more relaxed data consistency, i.e. by not using `NC_SHARE`. In this case, file header is synchronized among all processes in memories. No `MPI_File_sync()` will be called if header contents have changed. `MPI_File_sync()` will only be called when switching data mode, i.e `ncmpi_begin_indep_data()` and `ncmpi_end_indep_data()`.

A note about Large File Support

As of parallel-netcdf-0.9.2, we ship with support for CDF-2 formated data. With this format, even 32 bit platforms can create netcdf datasets (files) greater than 2GB in size. See the file README.large_files in the source tree for more information. CDF-2 also allows more special characters in the name strings of defined dimension, variables, and attributes.

The maintainers of the serial NetCDF library added support for the CDF-2 format in netcdf-3.6.0. The support was based largely on work from Greg Sjaardema.

The <u>CD-5</u> file format specification: supports 64bit data types and arrays with more than 2^{32} elements.

The CDF (or CDF-1) file format specification has been in use through netCDF library version 3.5.1.

File and Variable Limits

Both Parallel-netCDF and NetCDF share limitations on file and variable sizes. More information can be found on the FileLimits page.

Required Software

Parallel-NetCDF requires an MPI implementation with MPI-IO support. Most MPI libraries have this nowadays. A parallel file system would also go a long way towards achieving highest performance.

Related Projects

Parallel-NetCDF makes use of several other technologies.

- ROMIO, an implementation of MPI-IO, provides optimized collective and noncontiguous operations. It also provides an abstract interface for a large number of parallel file systems.
- One of those file systems ROMIO supports is PVFS, a high performance parallel filesystem for linux clusters.

Today, there are several options for high level I/O libraries. Here are some discussions on the role of Parallel-NetCDF in this ecosystem:

- pnetcdf_vs_hdf5?
- pnetcdf_vs_netcdf4?

Mailing List

We discuss the design and use of the Parallel-NetCDF library on the `parallel-netcdf@mcs.anl.gov` mailing list. Anyone interested in developing or using parallel-netcdf is encouraged to join. Visit the list information page for details.

The URL for the list archive is http://lists.mcs.anl.gov/pipermail/parallel-netcdf/. You can broswe even older mailing list messages at the older mailing list archives

Project Members

- Rob Latham, Rob Ross, and Rajeev Thakur (Argonne National Lab)
- Seung Woo Son, Wei-keng Liao, and Alok Choudhary (Northwestern University)
- Kui Gao (formally postdoc at Northwestern, now Dassault Systèmes Simulia Corp.)
- Jianwei Li (Northwestern, graduated in 2006)
- Bill Gropp (formerly ANL, now UIUC)

Citations

When referring to the Parallel-NetCDF project, please use our "permanent" URL: `www.mcs.anl.gov/parallel-netcdf`. The 'trac' or 'www-unix' URLs could change.

If you are looking for a reference to use in a published paper, please cite our SC2003 paper